

Missing data on ethnicity and socioeconomic status in high-impact medical journals

Sara Butterly

National Heart and Lung Institute, Imperial College London, London, United Kingdom

Keir Philip (✉ k.philip@imperial.ac.uk)

k.philip@imperial.ac.uk <https://orcid.org/0000-0001-9614-3580>

Saeed Alghamdi

National Heart and Lung Institute, Imperial College London, London, United Kingdom

Parris Williams

National Heart and Lung Institute, Imperial College London, London, United Kingdom

Jennifer Quint

National Heart and Lung Institute, Imperial College London, London, United Kingdom

Nicholas Hopkinson

National Heart and Lung Institute, Imperial College London, London, United Kingdom

Brief Communication

Keywords: ethnicity, socioeconomic status, health inequalities

Posted Date: November 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1019355/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Reporting participant ethnicity and socioeconomic status (SES) in clinical research is needed for interpretation and to inform discussion around health inequalities. We assessed the frequency of reporting of ethnicity (or 'race') and SES indicators in a sample of 100 research articles, in which participant level are reported, published in the 10 highest impact general medical journals in Spring 2021. 35 reported ethnicity and 13 SES, while 99 reported age, and 97 reported sex or gender. Among the articles not reporting ethnicity only 3(5%) highlighted this as a limitation, and only 6(7%) where SES data were missing. Median number of articles reporting ethnicity per journal was 2.5/10 (range 0-9). In conclusion, reporting of research participant ethnicity and socioeconomic status in high-impact medical journals remains poor, and this omission is rarely acknowledged as a limitation. This situation persists despite the well-established importance of this issue. Standardized explicit minimum standards are required.

Main Text

Information about the ethnicity and socioeconomic status of participants in clinical research is needed for the interpretation, generalisability and pooling of data as well as to inform discussion around health inequalities. The relevance of ethnicity and socio-economic status to health and biomedical research is well established but has been emphasised by the COVID-19 pandemic, during which specific ethnic groups and poorer individuals have been disproportionately affected¹. The causal pathways driving health disparities are complex and multifactorial, however under-reporting of participant characteristics has been identified as a potential contributory factor²⁻⁴.

The International Committee of Medical Journal Editors (ICMJE) recommendations⁵, and some journal instructions to authors promote inclusion of these data^{6,7}. Previous studies have identified that reporting is frequently incomplete with limited progress made over the last three decades⁸⁻¹³. Recent years have seen an increased focus on ethnicity and socioeconomic status in medicine, however there is a lack of research as to whether this has resulted in better reporting.

To evaluate the current situation in this area, we assessed the frequency of reporting of ethnicity (or 'race') and socioeconomic status indicators in a sample of research articles published in high impact general medical journals in Spring 2021.

We identified the 10 highest ranked journals as per Google scholar 'Health and Medical (general)' category up to April 2021. At the time of data collection these were The New England Journal of Medicine (NEJM), The Lancet, the Journal of the American Medical Association (JAMA), Proceedings of the National Academy of Sciences of the United States of America (PNAS), Nature Medicine, Public Library of Science One (PLOS One), The British Medical Journal (BMJ), Cochrane, Cell Metabolism, and Science Translational Medicine. PNAS and PLOS One include a wide range of subject areas therefore the subsections 'Biological Sciences, Medical Science' and 'Clinical Medicine' were used respectively. From

each of these 10 journals, we selected the 10 most recent journal articles that report participant level data. Laboratory studies using human derived tissues or cells were included if donor information was provided. Journal reporting guidance and requirements were also assessed by evaluating author guidelines, websites, and contacting the respective editorial/publishing teams. Data were collected on which participant level characteristics were reported and how. Data collection and analysis was conducted by SCB, KEJP, SMA and PW. All papers were reviewed independently by at least two researchers.

Ethnicity and race are related yet different constructs and arguably the latter term should be abandoned¹⁴. However, given the frequent lack of standardisation in the literature and that the terms are in practice often used interchangeably we accepted the use of either term. Similarly, regarding reporting of socioeconomic status indicators, various often inconsistent methods are used, therefore we opted to assess both direct measures such as the Index of Multiple Deprivation, but also measures from which socioeconomic status could be inferred such as educational attainment and job role. The focus being if, rather than how, such measures are reported.

650 publications were assessed to identify 100 meeting inclusion criteria (figure 1). Of one hundred research articles included, 35 reported ethnicity (or race) and 13 reported socioeconomic status. By contrast, 99 reported age, and 97 reported sex or gender (Table1). Among the articles not reporting ethnicity only 3 (5%) highlighted this as a limitation, and only 6 (7%) highlighted where socioeconomic status data were missing. Median number of articles reporting ethnicity per journal was 2.5/10 (range 0/10 (PLOS One), to 9/10 (JAMA)). Only 2 journals explicitly requested reporting of participant ethnicity (or race), and 1 requested socioeconomic status. Types of research included – interventional studies (n=30), cohort studies (n=35), case-control studies (n=3), systematic reviews and meta-analyses (n=16), epidemiological and surveys (n=3), and other (n=13). Twenty of the 100 were laboratory studies (either observational or involving interventional manipulation of samples) using human samples, of which 4 reported ethnicities of sample donors (of others, none mentioned as a limitation), and none reported socioeconomic status.

Among the 24 papers describing clinical trials, 50% reported ethnicity, with none highlighting the absence of these data as a limitation. 12.5% of trials reported an indicator of socioeconomic status, with one of the 21 not reporting socioeconomic status highlighting this absence as a limitation.

Of note, two of the research articles included in our sample identified ethnicity as being relevant to their research topic, yet did not provide relevant data on their study participants or highlight the lack of this data as a limitation of their study '*in the case of DNA-based mutation testing, poor sensitivity in detecting mutations in infants from ethnic and racial minority groups*'¹⁵, and '*peripheral oxygen saturation can substantially differ from the SaO₂ under certain conditions and may be less accurate in Black patients than in White patients*'.¹⁶.

Figure 1: Consort of included/excluded articles (see end of manuscript)

Table 1: Reporting of race or ethnicity and socioeconomic status (see end of manuscript)

The majority of research published in high-impact medical journals does not include data on the ethnicity and socioeconomic status of participants, and this omission is rarely acknowledged as a limitation. This finding echoes related historical research,⁸⁻¹³ but its persistence is of concern and is surprising given current awareness of such issues^{17,18}.

These findings have important implications for the interpretation and application of research findings, both within academia and beyond, with the ongoing omission no longer justifiable as simple oversight. As highlighted by Baker et al.¹⁹ in relation to data relating to LGBTQI+ communities, but equally relevant here, '*Data are fundamentally political: decisions about which data are collected and which are overlooked both reflect and shape policy and program priorities.*'

Our results could have multiple contributory factors. For some research including secondary data analyses, ethnicity and socioeconomic status data may not have been available to the researchers, but given the lack of explanation, it remains unclear if these data were unavailable, or available but not included in publications. The low level of reporting in controlled clinical trials suggests issues beyond unavailability of data, as in these studies such data would be simple to collect. Additionally, given research successfully reporting these data, the justification for these omissions remains unexplained.

The increased frequency of reporting ethnicity compared to socioeconomic status, may indicate differences between the perceived relevance of these variables. This would be in keeping journal author guidelines and ICMJE recommendations that encourage the inclusion of relevant demographic variables to ensure representative samples⁵, more often explicitly stating race and/or ethnicity, than socioeconomic status. The relevance of these factors may not have been apparent to authors and editorial teams, however ICMJE Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals⁵ states '*Because the relevance of such variables as age, sex, or ethnicity is not always known at the time of study design, researchers should aim for inclusion of representative populations into all study types and at a minimum provide descriptive data for these and other relevant demographic variables.*' Of note, not all of the journals in our sample state that they follow the ICMJE recommendations²⁰. However, whether or not the journal states they follow guidance or not, this has no impact upon the relevance of these data and the importance of reporting them. Additionally, Maduka et al²¹ found no difference between journals stating they follow ICMJE recommendations, and those that do not, in the frequency of reporting race and ethnicity in a sample of surgical research publications in 2019.

Certain considerations require highlighting. Firstly, different approaches to selecting research papers may alter findings. Secondly, we identified high-impact journals using the google scholar H5 index but acknowledge various other equally valid methods exist. Thirdly our analysis focused on if ethnicity and/or race was reported, but we acknowledge that these are not synonymous terms. In addition to *if* these variables are reported, *how* they are reported is also an important area for discussion and research. The widespread omissions identified by this research suggests a structural problem. Indeed, we the authors have published research which would have met the inclusion criteria and failed to report these specific characteristics. Our intention is to highlight an issue and suggest approaches to address it.

Given that inadequate reporting persists despite research highlighting the issue, author and ICMJE recommendations, and the current socio-political climate, there is a clear need for more explicit requirements that are adhered to in practice. This is likely best achieved if steps are integrated into each stage of the research process, from protocol to publication. For example, Fain et al²² compared reporting of race and ethnicity on ClinicalTrials.gov before and after the requirement to report these data (if collected), was introduced, finding that this was associated with an increase from 42% to 92%. Similar explicit requirements could be taken in EQUATOR guidelines²³, and research ethics applications. From our sample, the journal JAMA had the most explicit guidance for reporting race and ethnicity, and this variable was reported in 9/10 of the articles we reviewed. Of note from 2022 the *New England Journal of Medicine* will be requiring authors of research articles to provide data on the representativeness of the sample including race or ethnic group²⁴, though it is unclear if socioeconomic status indicators will also be required.

The reporting of ethnicity and socioeconomic status in high-impact medical research remains poor, despite a consensus on its importance. Omission of these participant characteristics limits the interpretation, generalisability, and pooling of data, that are required to facilitate informed discussion around health inequalities. Guidance and encouragement have so far proven insufficient to change practice in this area. Standardised, explicit, minimum standards are required.

Declarations

Contributors:

SCB, had the original idea for the study. SCB, KEJP, SMA and PW collected the data. All authors (SCB, KEJP, SMA, PW, JKQ and NSH) contributed to the design of the study. KEJP analysed the data initially, which was verified by SCB, SMA and PW. KEJP wrote the first draft of the manuscript. All authors (SCB, KEJP, SMA, PW, JKQ and NSH) critically appraised the manuscript and approved it for submission, and had full access to the data and can take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author attests that all listed authors (SCB, KEJP, SMA, PW, JKQ and NSH) meet authorship criteria and that no others meeting the criteria have been omitted.

Funding:

KEJP is supported was supported by the Imperial College Clinician Investigator Scholarship (no specific grant number/code). The funders had no say in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests:

None reported.

Data sharing:

All data used in this study are publicly available.

Tables

Table 1: Reporting of ethnicity and/or race, and Socioeconomic Status indicators in research articles

	N	Additional notes
Report		
participant level	100	
characteristics		
Report:		
Report	35/100 report	Range per journal: JAMA 9/10, with clear guidance that this information is expected.
ethnicity and/or race	65 Not report	
Noted limitations	in 62 of the 65 do not state this as a limitation 3 Do highlight this as a limitation.	Some studies identify race and ethnicity as being relevant to the research focus, yet did not provide relevant data on their study participants or highlight this a limitation of their study e.g. <ul style="list-style-type: none"> <i>'in the case of DNA-based mutation testing, poor sensitivity in detecting mutations in infants from ethnic and racial minority groups'</i> (DOI: 10.1126/scitranslmed.abd8109) <i>'peripheral oxygen saturation can substantially differ from the SaO₂ under certain conditions and may be less accurate in Black patients than in White patients.'</i> (DOI: 10.1056/NEJMoa2032510)
Socioeconomic status indicator reported	13/100 report at a measure of SES (6 direct measure e.g. Index of Deprivation, Poverty income ratio; 7 measures from which SES can be inferred eg educational attainment, job role)	

87/100 did not report any indication
of SES

Noted in 6/87 identified this as a limitation
limitations

Age reported 99/100

Sex or Gender 97/100
reported

Percentages not given as most results have 100 as the denominator.

References

1. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.
2. Chastain DB, Osae SP, Henao-Martinez AF, Franco-Paredes C, Chastain JS, Young HN. Racial Disproportionality in Covid Clinical Trials. *The New England journal of medicine* 2020;383:e59.
3. Krieger N, Waterman PD, Chen JT, Testa C, Hanage WP. Missing again: US racial and ethnic data for COVID-19 vaccination. *Lancet* 2021;397:1259–60.
4. Webb Hooper M, Napoles AM, Perez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *Jama* 2020;323:2466–7.
5. ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. Updated December 20192019.
6. Information for Authors. 2021. at <https://marlin-prod.literatumonline.com/pb-assets/Lancet/authors/tl-info-for-authors.pdf>.)
7. Instructions for Authors. 2021. at <https://jamanetwork.com/journals/jama/pages/instructions-for-authors>.)
8. Geller SE, Koch A, Pellettieri B, Carnes M. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? *J Womens Health (Larchmt)* 2011;20:315–20.
9. Brahan D, Bauchner H. Changes in reporting of race/ethnicity, socioeconomic status, gender, and age over 10 years. *Pediatrics* 2005;115:e163-6.
10. Kanakamedala P, Haga SB. Characterization of clinical study populations by race and ethnicity in biomedical literature. *Ethn Dis* 2012;22:96–101.
11. Bokor-Billmann T, Langan EA, Billmann F. The reporting of race and/or ethnicity in the medical literature: a retrospective bibliometric analysis confirmed room for improvement. *Journal of clinical epidemiology* 2020;119:1–6.

12. Lee SJ, Kavanaugh A. A need for greater reporting of socioeconomic status and race in clinical trials. *Ann Rheum Dis* 2004;63:1700–1.
13. Ma IW, Khan NA, Kang A, Zalunardo N, Palepu A. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *Journal of clinical epidemiology* 2007;60:572–8.
14. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the Use of Race in Medical Research. *Jama* 2021;325:623–4.
15. Ray TR, Ivanovic M, Curtis PM, et al. Soft, skin-interfaced sweat stickers for cystic fibrosis diagnosis and management. *Sci Transl Med* 2021;13.
16. Schjorring OL, Klitgaard TL, Perner A, et al. Lower or Higher Oxygenation Targets for Acute Hypoxemic Respiratory Failure. *The New England journal of medicine* 2021;384:1301–11.
17. Flanagin A, Frey T, Christiansen SL, Bauchner H. The Reporting of Race and Ethnicity in Medical and Science Journals: Comments Invited. *Jama* 2021;325:1049–52.
18. Race and Medicine. 2021.
19. Baker KE, Streed CG, Jr., Durso LE. Ensuring That LGBTQI+ People Count - Collecting Data on Sexual Orientation, Gender Identity, and Intersex Status. *The New England journal of medicine* 2021;384:1184–6.
20. Journals stating that they follow the ICMJE Recommendations. 2021. at [http://www.icmje.org/journals-following-the-icmje-recommendations/.\)](http://www.icmje.org/journals-following-the-icmje-recommendations/.)
21. Maduka RC, Broderick M, White EM, et al. The Reporting of Race and Ethnicity in Surgery Literature. *JAMA Surg* 2021.
22. Fain KM, Nelson JT, Tse T, Williams RJ. Race and ethnicity reporting for clinical trials in ClinicalTrials.gov and publications. *Contemp Clin Trials* 2021;101:106237.
23. Reporting guidelines. 2021. (Accessed 04/10/2021, 2021, at [https://www.equator-network.org/reporting-guidelines/.\)](https://www.equator-network.org/reporting-guidelines/.)
24. Rubin E. Striving for Diversity in Research Studies. *The New England journal of medicine* 2021.

Figures

Journals with the highest Google Scholar H5 index: NEJM, The Lancet, JAMA, PNAS, Nature Medicine, PLOS One, BMJ, Cochrane, Cell Metabolism, and Science Translational Medicine

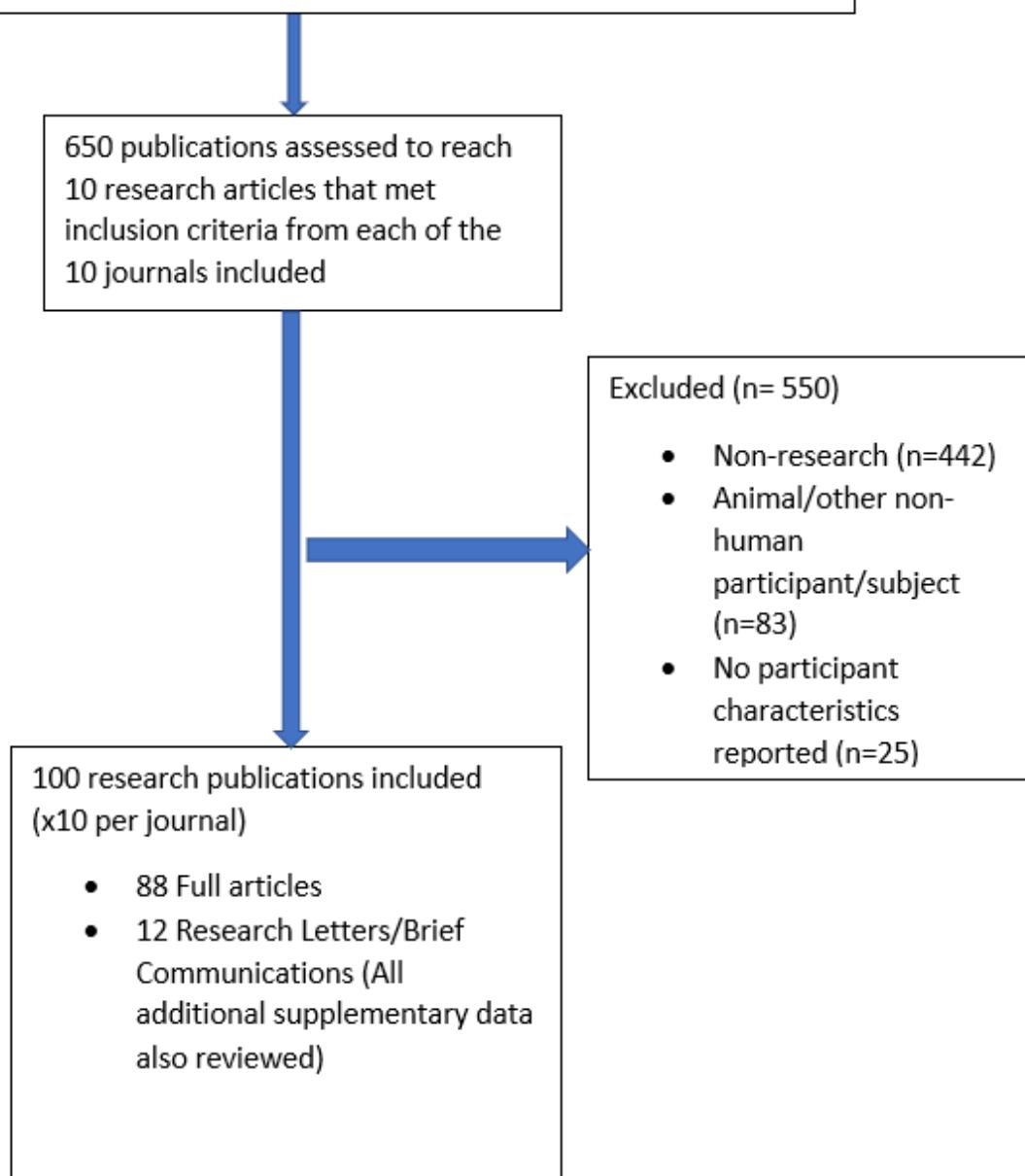


Figure 1

Consort diagram of study inclusion/exclusion